Crowd Analysis with Target Tracking, K-means Clustering and Hidden Markov Models

Maria Andersson, Joakim Rydell
Div of Sensor Informatics, Dept of Sensor & EW Systems
Swedish Defence Research Agency
SE-581 11, Linköping, Sweden
maria.andersson@foi.se, joakim.rydell@foi.se

Louis St-Laurent [1], Donald Prévost [1], Fredrik Gustafsson[2]
[1] INO, Québec, Canada
[2] Dept. of Electrical Engineering, Linköping University,
SE-581 83, Linköping, Sweden
louis.st-laurent@ino.ca, donald.prevost@ino.ca,
fredrik@isy.liu.se

***Abstract*—The paper presents a framework for crowd analysis that can handle both sparse and dense crowds, by combining micro- and macroscopic crowd analysis approaches. The paper focuses on detection, tracking and behaviour of dense crowds. We use multiple target tracking (MTT), group tracking, *K*-means clustering and hidden Markov models (HMM). *K*-means clustering is used to decide if micro- or macroscopic approaches should be used. A first evaluation, based on recorded and simulated data sets, has been done. The evaluation shows that MTT works well when the crowd is relatively sparse. When the crowd becomes dense track identities are easily switched between tracks. For dense crowds centroid-based group tracking is proposed. The algorithms for dense crowd detection and behavior recognition show promising results. The accuracies of the algorithms range from 84 % and above. Increased internal crowd activities will, however, temporarily reduce the accuracy of the centroid-based group tracking.**

Keywords-component; crowd analysis, crowd behavior, multiple target tracking, group trackig, hidden Markov models, K-means clustering.

I. INTRODUCTION

During the last years crowd analysis has been applied in various different applications, including visual surveillance, crowd management and public space design. In visual surveillance crowd analysis is used for automatic detection of anomalies or threatening events. In crowd management crowd analysis is used to analyze situations such as sport events, large concerts and public demonstrations, in order to avoid crowd related disasters. For public space design crowd analysis is used to provide guidelines for the design of public spaces [1].

In this paper we focus on the first application, i.e. to use crowd analysis to improve situational awareness in visual surveillance. Automatic crowd analysis can considerably improve the possibilities for an operator to detect, at an early stage, important events in the often very large amount of information. With automatic crowd analysis it is possible to foresee different states of the crowd, for example; crowd size, crowd density, crowd flow, crowd speed and crowd anomaly (e.g. riots and chaotic acts).

Crowd analysis can also be used to localize abnormal regions in crowded and complex scenes, where high resolution analysis is done in a next step. Such strategy will minimize the number of computations, while still obtaining detailed results.

Models for crowd analysis are based on characteristic crowd features and the relationships between them. Typical crowd features include; detection, colour, shape, texture, motion, size, velocity and sound level.

Crowd analysis can be divided into three approaches [1]. In the microscopic approach the people are analyzed as discrete individuals. This information is then used to build up knowledge on the crowd. In the macroscopic approach the crowd is analyzed as a whole, without information on the individuals. The third approach is a combination of the two, which can further be divided according to:

1. Keeping a crowd as a homogeneous mass but considering an internal force.
2. Keeping the characters of the persons while maintaining a general view of the entire crowd.

Crowd models are formulated in many different ways. In for example [2] the authors are inspired by theories from gas kinetics and fluid dynamics in the modeling of crowd behavior. Reference [3] uses the social force model, in a macroscopic approach, to describe the crowd. In the social force model interactions in the crowd are related to personal motivations and environmental constraints. For example, people try to not collide with each other. People that know each other tend to walk close to each other. People that do not know each other tend to have a certain distance to other people. Buildings and stationary objects force the people to move in certain directions.

A combination of micro- and macroscopic approaches is proposed in [4] where the purpose is to understand group behavior in subway stations. Detection and tracking of individuals together with group tracking form a basis for group behavior analysis. Group tracking can often give better performance in crowded environments, compared to traditional MTT. For example, [5] suggests group tracking with crowd

models that are based on evolving graphs structures. Other aspects of group tracking are discussed in [6] and [7].

For sparse crowds, the tracking of individuals will give a good basis for the overall situation assessment. For example, [8] presents a pedestrian tracking method that includes informative descriptions of the road structure. This information (including road-constrained tracking and unconstrained tracking) is used to enhance tracking performance.

A. Objective

The objective of this paper is to present the preliminary framework for automatic crowd analysis in urban environments where both sparse and dense crowds can be analyzed. In the literature it is not common to study a combination of the two; instead the focus is often on one of the situations.

In earlier work we have focused on dense crowd analysis from a macroscopic point of view. In [9] and [10] we used optical flow and foreground pixels to derive features for the crowd as a whole, and for analyzing crowd behavior. In this paper we will start to investigate how detection and tracking of individuals can contribute to crowd analysis and be combined with a macroscopic approach.

B. Outline

The paper is organized as follows. Section II introduces the idea of the crowd analysis framework. Section III presents briefly the background theories. Section IV explains the application of the theories on the crowd analysis framework. Section V shows experiments and results, where we have implemented a part of the general framework. Section VI presents a summary of the results together with some conclusions. Finally, section VII discusses future work.

II. Crowd Analysis Framework

The framework uses a combination of micro- and macroscopic approaches. The procedure is briefly illustrated in Fig. 1. The link between the approaches is found in the definition of a dense crowd, which in this framework is given from clustering analysis.

When a dense crowd is detected the macroscopic approach should be used for tracking and behavior analysis. The information given from the macroscopic approach will describe the dense crowd as a whole. Less (or possibly no) information will be obtained from individuals in the dense crowd.

In sub areas where no dense crowds exist, the microscopic approach should be used for tracking and behavior analysis. In such areas quite detailed information on specific persons can be obtained.

Also the behavior of dense crowds is analyzed. The behavior is either calm or active. In a calm crowd people stand still or move closely together in the same direction, with similar and low speed. In an active crowd people move randomly, in different directions and speed. Also high speed is associated with an active crowd. An active crowd can be stationary or move. The active crowd can represent for example fights and/or riots.

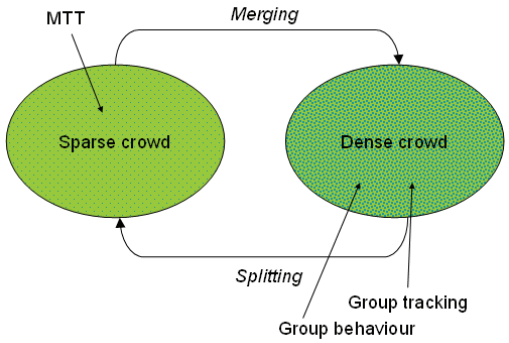

Figure 1. The crowd analysis framework where both dense and sparse crowds can be analyzed.

We define sparse and dense crowds as follows:

- In a sparse crowd there is enough space between the people so that each individual can be detected and tracked using traditional methods such as MTT.
- In a dense crowd the people are close to each other and occlusion is frequent. The detection and data association processes becomes much more uncertain.

III. Background Theory

A. Multiple Target Tracking (MTT)

MTT is used to localize the person in space and time in a scene that consists of several persons. Also the velocity information can be derived. A crucial part for MTT is data association, i.e. the process to recognize the same person, among other persons, in consecutive frames. Typical techniques for MTT include Kalman filtering, extended Kalman filtering (EKF) and particle filtering. Typical techniques for data association include global nearest neighbor and probabilistic data association [11].

Another crucial part is the detection of persons. It is important that the detection can be done in a robust way in order to obtain, in the next step, robust tracking. Typically, detection algorithms are based on foreground-background segmentation.

Crowded scenes increase the complexity for MTT, especially in the detection and data association processes. The reason is that there are multiple, and closely, moving persons which will temporarily be hidden by other persons/objects.

B. Group Tracking

In crowded scenes detections may not be received from all persons in all frames because of occlusion. Fewer tracks may be present than the actual number of persons. Moreover, the tracks may easily switch identities. For these scenes group tracking is often a better alternative. Group tracking can be divided into three approaches [11], which are:

1. Group tracking without individual tracks. This is the simplest approach with minimum processing load.
2. Group tracking with simplified individual tracks. This is a more complex logic leading to increased processing load, but with potentially more accurate group tracks.
3. Individual target tracking which is supplemented by group tracking. This alternative results in maximum processing load but with the potential for the most accurate tracking performance.

Group tracking uses the same processes as the conventional tracking methods, such as detection, association and prediction. An additional process is to update the knowledge of the size and form of the crowd.

C. Cluster Analysis

Cluster analysis is used for segmenting a collection of objects into clusters. Within each cluster the members are more closely related to each other than to cluster members who belong to other clusters [12].

K-means clustering is one of the most popular iterative clustering methods. The similarity measure, that is used to estimate the closeness of objects, is the squared Euclidean distance.

Output data from *K*-means clustering include the centroid coordinates, the number of cluster members and the sum of the distances between each cluster member and the corresponding centroid.

D. Motion Recognition

In motion recognition the purpose is to recognize specific motion patterns that are more or less hidden in often large sets of data. An often used model for motion recognition is the HMM [13].

In brief, HMM consists of two stochastic processes. The underlying (hidden) process can not be observed directly, but can be observed indirectly through the second stochastic process which produces sequences of observations. The states of the HMM represent some unobservable condition of the system.

The HMM is defined by the parameters A, B, π, N and M. A is the probability distribution of the state transitions, B is the probability distribution of the observations in each state and π is the probability distribution of the initial state. N is the number of states and M is the number of observations. The parameters can be obtained by training the HMM on a set of relevant training data.

IV. The Formulation of the Crowd Analysis Framework

A. Tracking Persons in Sparse Crowds

Since the cameras are static in the scenarios it can be expected that the scene background will change relatively slowly. We detect people using foreground-background segmentation, where each pixel in the background is modelled as a Gaussian distribution. A pixel is considered to belong to a foreground object if its value does not fit the model. The background model is updated continuously, thereby adapting to slow changes in lighting conditions etc. Persons are detected by inspecting clusters of foreground pixels, looking for human-like shapes.

Detected persons are tracked using a MTT algorithm that is based on EKF. In each video frame, all detections are matched against the currently tracked persons. Observations which are associated to an existing track are used to update its filter estimate, while new tracks are created based on non-associated observations. We currently use global nearest neighbour association.

The tracking provides trajectories for the people moving in the scene. These trajectories can be used for detecting specific events, such as persons merging into a dense crowd, persons staying in the scene for unusually long times, persons walking in the wrong direction, or running persons. It is also possible to detect more complex behaviors, such as interactions between multiple persons.

B. Tracking Persons and Crowds

1) Detection in Raw Images

Foreground-background segmentation provides a set of detected persons represented with ellipsoids. Rectangles or any other shape are of course possible, but we stick to ellipsoids for simplifying the mathematical fusion framework. Each ellipsoid is represented with a center coordinate z_k^p in the image plane, and "covariance" Σ_k^p, for $k = 1, 2, \ldots, K$, with K detections. Here, superindex 'p' stands for persons. We will not enforce any structure on Σ_k^p, though $\mathrm{diag}(\sigma_x^2, \sigma_y^2)$ with $\sigma_y \approx 4\sigma_x$ is one common choice for upright persons.

2) Merging Persons to Crowds

We apply a *K*-means clustering algorithm to find candidates for clusters of persons that can be treated as one crowd. The output from the clustering algorithm is, among other things, a cluster center z_l^c in the image plane. Here, superindex 'c' stands for cluster to distinguish it from persons. The cluster l consists of the persons $k \in K_l$, where K_l defines a set of indices of persons. The cluster center and its covariance (representing an area covering the persons) are computed using standard merging formulas as

$$z_l^c = \frac{1}{|K_l|} \sum_{k \in K_l} z_k^p , \tag{1}$$

$$\Sigma_l^c = \frac{1}{|K_l|} \sum_{k \in K_l} \Sigma_k^p + \left(z_k^p - z_l^c\right)\left(z_k^p - z_l^c\right)^T . \tag{2}$$

Here, $|K_l|$ denotes the cardinality of the set K_l that is the number of persons in the set. To decide how dense a cluster of persons is we propose, as a general approach, to use the property that $\det(\Sigma)$ is proportional to the area of the ellipsoid. The area of the cluster normalized with the total area of all

persons included in the cluster is a good indicator of crowd density. Therefore an appropriate dense-crowd measure is:

$$d_l = \frac{\det(\Sigma_l^c)}{\Sigma_{k \in K_l} \det(\Sigma_k^p)}. \quad (3)$$

For a cluster with only one person, we get $d_l = 1$. For a cluster with $|K_l|$ persons on the same spot, $d_l = 1/|K_l|$. When the persons are moved away from each other, d_l increases indefinitely. That is, a small d_l (much less than one) indicates a crowd. A thresholding procedure can thus be applied,

$$d_l < D. \quad (4)$$

The output from the clustering algorithm is thus a set of validated crowds z_l^c, $l = 1, 2, \ldots, L$ and a set of persons z_k^p, $k = 1, \ldots, K$ that do not belong to a crowd.

In the preliminary framework presented here we have implemented a simplified version the dense-cluster measure, which is denoted d_l^*. d_l^* is based on the average value of the Euclidean distances between z_l^c and the corresponding cluster members. In the next version of the framework we will implement the dense-crowd measure as (3).

3) Tracking of Persons and Crowds

Tracking is basically done in the same way for persons and crowds. The state consists of at least position and velocity in standard motion models. Position is still given in the same plane. We use the simplest motion model with only position and velocity in the state vector. The total model is the linear, and the Kalman filter applies for tracking. There are in total $L + K$ Kalman filters running in parallel for all clusters and unclustered persons.

4) Multiple Target Tracking

The Kalman filter relies on a correct association of persons z_k^p and crowd z_l^c centers at each time. Association is performed using a simple nearest neighbour approach, where the predicted center z_k^c and covariance Σ_k^c from each filter are compared to the outputs from the clustering algorithm. More complex algorithms like the Hungarian (or auction) algorithm are possible, but not needed in the scenarios we have studied.

C. Estimation of Crowd Behavior

The dense-crowd measure d_l^* is used as input data to the crowd behavior analysis. Changes in d_l^* over time serve as input data to HMMs which estimates the degree of internal crowd activity. Two HMMs have been modeled, i.e.

- λ_1 = calm crowd, where the internal motions (or forces) are relatively constant and often slow.
- λ_2 = active crowd, where the internal motions (or forces) are continuously changing and often fast.

For λ_1 and λ_2 we use $N = 3$ and $M = 3$. Artificial training data was used to train λ_1 and λ_2, i.e. to obtain relevant model parameters. The training was done with the Expectation-Maximization algorithm [14]. The likelihood L for an observation sequence $O = O_1, O_2, \ldots, O_t$ is given by the following:

$$L = (P(O \mid \lambda)) \quad (5)$$

If $\lambda = \lambda_1$ then (5) estimates the likelihood of an observation sequence coming from a calm crowd. If $\lambda = \lambda_2$ then (5) estimates the likelihood of an observation sequence coming from an active crowd.

Equation (5) is preferably calculated with the forward-backward algorithm [13]:

$$\alpha_t(i) = P(O_1, O_2 \ldots O_t = S_i \mid \lambda) \quad (6)$$

which describes the probability of the partial observation sequence $O_1, O_2, \ldots, O_t$ and state S_i at time t, given the model λ.

The observation symbols O_t can take the values 1, 2 or 3, according to the following relations:

$$d_{l,t}^* = d_{l,t-1}^* \quad (7)$$

$$d_{l,t}^* < d_{l,t-1}^* \quad (8)$$

$$d_{l,t}^* > d_{l,t-1}^* \quad (9)$$

Equation (7) means that the crowd density at t is the same compared to the density at t - 1. If this is the case $O_t = 1$. Equation (8) means that the crowd density has increased at t, compared to the density at t - *1*. If this is valid $O_t = 2$. Equation (9) means that the crowd density has decreased at t, compared to the density at t - *1*. If this is valid $O_t = 3$.

D. Performance Measure for the Evaluations

In Section V the algorithms for dense crowd detection and behavior analysis are evaluated against recorded data. The performance of the algorithms $\eta_{scenario}$ is estimated according to:

$$\eta_{scenario} = 100 \times \left(\frac{D_{corr}}{D_{tot}} \right) \quad (10)$$

where D_{corr} represents the number of correct decisions and D_{tot} represents the true decisions.

V. Experiments

The data sets that are used in Sections V.A – V.B can also be studied in [15].

A. People Merge to a Dense Crowd

In this experiment there are nine people moving in a parking place. Initially they come from different directions. They approach each other and form a dense, calm and stationary crowd in the middle of the scene. After a while the dense crowd splits and people leave the scene in different directions. The scenario is observed by two visual cameras and two thermal infrared cameras. The field of view of the cameras can be seen in Fig. 2. In this experiment we investigate the performance of MTT and the detection of dense crowds.

1) Tracking in Sparse Crowds

Fig. 2 shows a snapshot of the detection and tracking of the persons. The white numbers correspond to different track identification numbers and the tails behind each person show their locations during the last few seconds.

The current detector and tracker work well as long as persons in the scene are relatively well separated. When dense groups are formed, erroneous associations tend to cause persons to switch track identification numbers.

Ongoing work aims at solving this problem by using visual attributes such as colour histograms to facilitate the association process, and by using multiple-hypothesis tracking.

2) Detection of Dense Crowds

The merging of people into a dense crowd is shown in Fig. 3. In Fig. 4 the results from the dense-crowd detection are shown. The method searches for a cluster that has $d_l^* < 2$ (m). Input data is assumed to be correct detections of the persons.

A crowd of at least four people is detected at *Time* = 9 s. Since more people enter the scene during the existence of the dense crowd, but are still far from the crowd, d_l^* temporarily increases above the threshold. To reduce this effect we introduce a time rule.

Figure 2. A snapshot of the detection and tracking. Top row: visual images, bottom row: thermal infrared images.

Figure 3. A dense crowd is formed (here seen from a thermal infrared camera). To the right the gate around the dense crowd is shown.

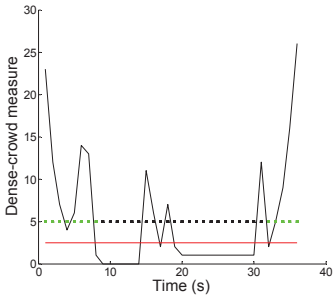

Figure 4. Crowd state: dense or sparse crowd. The *y* axis shows the dense-crowd measure d_l^*. The dense crowd exists for $9 < Time < 31$. When the black curve is below the red line the people in the scene stand or move close to each other. The dots show true the decisions (i.e. dense or sparse). The black dots represent the dense crowd. The green dots represent the sparse crowd.

With this rule the method will keep in mind the existence of a dense crowd despite that people continue to approach the crowd. In this paper, if a dense crowd exists for at least 3 s then the group should still be defined as a dense crowd even if more people are entering the scene. The dense crowd splits at *Time* = 31 and d_l^* increases again.

The accuracy of the crowd detection algorithm in this scenario is $\eta_{scenario} = 88$ %.

B. People Merge to a Dense and Active Crowd

In this scenario there are seven people that merge into a dense crowd which exists for 25 s. During this time there are fights between several people. After the fights the crowd splits and most people leave the scene. Fig. 5 shows snapshots of the fights as observed from the two visual cameras. In this scenario we investigate the detection of dense crowd and the crowd behavior. Input data are assumed to be correct detections of persons.

1) Detection of Dense Crowd

A crowd is formed at *Time* = 4 s. In the beginning there are more and more people connecting with the crowd and it has found a more final form at *Time* = 6 s. At *Time* = 15 s the fights start which can be observed with an increase of d_l^*, as the people temporarily move away from each other during the fight. The crowd splits at *Time* = 22 s and at *Time* = 23 s the number of people is less than four. Fig. 6 shows the results from the dense-crowd detection. The performance in this scenario is $\eta_{scenario} = 84$ %.

Figure 5. The two images illustrate the dynamics of the crowd, where the time difference between the two images is 1 s.

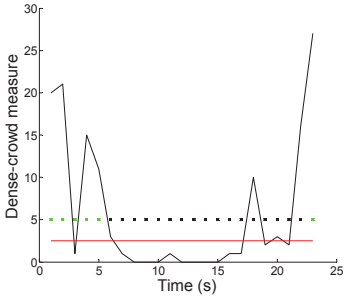

Figure 6. Crowd state: dense or sparse crowd. The *y* axis shows the dense-crowd measure d_l^*. The fights take place when 15 < *Time* < 23 s. The dots show true the decisions (i.e. dense or sparse). The black dots represent the dense crowd. The green dots represent the sparse crowd.

2) Estimation of Crowd Behavior

The changes of d_l^* in consecutive time steps indicate internal activities, or forces, in the crowd. This information is used to estimate the degree of the internal activities, according to (7-9). The result of the crowd behavior analysis is presented in Fig. 7. The blue curve represents λ_1 (calm crowd) and the magenta curve represents λ_2 (active crowd). The HMM with the highest log-likelihood should be the basis for the crowd behavior decision at each point of time. When the dense crowd is detected the observation sequence will have only one observation, i.e. $O = O_1$. When *Time* = 10 s O finally contains all the observations and the behavior analysis can be done completely. This is the reason why the log-likelihood of both λ_1 and λ_2 are increasing in the beginning.

The crowd is calm until the fights start at *Time* = 15 s. From that time the log-likelihood for calm crowd is reducing and at the same time the log-likelihood for active crowd is increasing. At *Time* > 21 s the movements become strongly irregular, which is also obvious in Fig. 7. The fights end at *Time* = 23 s. At *Time* > 23 s there are only three people in the scene which is a too small crowd to be analyzed as a dense crowd in this case. For the crowd behavior $\eta_{scenario}$ = 84 %.

C. Dense Crowd in Motion

This experiment shows a dense and calm crowd that moves a short time period in the scene. The crowd is observed by one visual camera.

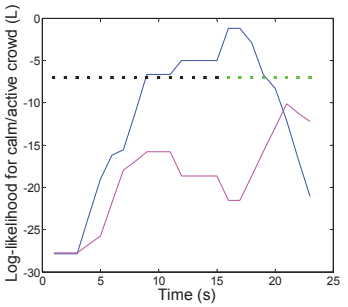

Figure 7. The log-likelihood for a calm crowd (blue) and the log-likelihood for an active crowd (magenta). True states are represented by the dots. Black dots represent calm crowd. Green dots represent active crowd.

Figure 8. A calm crowd in motion.

A snapshot from the experiment is shown in Fig. 8. Input data are assumed to be correct detections of the persons. The result from the crowd behavior analysis is presented in Fig. 9, with the log-likelihood for calm and active crowds according to (7–9). In the middle of the scenario some people reduce their velocities in relation to the others. The change in velocity will cause a change in cluster area and d_l^*. At this time there is a short decrease in log-likelihood for a calm crowd, which can be seen at approximately the half distance in Fig. 9. Soon these people will catch up with the others, and the log-likelihood for calm crowd is increased again. For the crowd behavior in this scenario, $\eta_{scenario}$ = 100 %. However, the result indicates that when people just move irregularly for a shorter time period (without being anomaly or threatening in some way), there may be a small influence on the crowd behavior and consequently a risk for the wrong decision concerning the state of the crowd.

D. The Influence of Internal Crowd Activities on group tracking performance

If there are intense activities within the crowd (from e.g. fights, violence), d_l^* can be expected to vary strongly. The purpose of this experiment is to investigate, in this case with simulated data, the possible effects of intense activities on group tracking performance. The scenario is as follows: a calm and dense crowd is moving in the scene where all people initially have similar velocities.

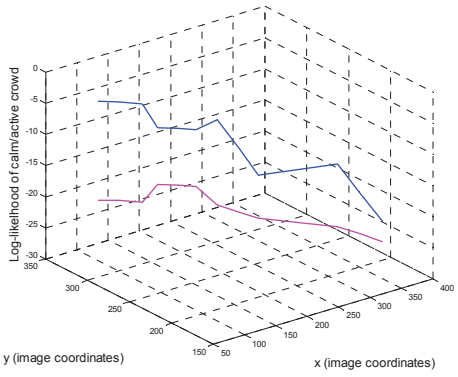

Figure 9. A small and calm crowd moves in the scene. The blue curve represents a calm crowd and the magenta curve represents an active crowd. During the whole scenario the dense crowd is, according to HMM, expected to be calm. However, some irregular but normal motion changes, by some people, will cause a small drop in log-likelihood for calm crowd.

During a time period a few people start to move in an irregular manner with random changes in speed and direction. Fig. 10 shows the log-likelihood for calm and active crowds. As can be seen, when irregular motions are introduced for 30 < *Time* < 40 there is a change from calm crowd to active crowd. Fig. 11 shows the position estimates from the group tracking (black curve), where the measurement data are z_l^c from the *K*-means clustering. The red curve shows z_l^c. When 30 < *Time* < 40 the position error is increased considerably, as a result of the increased activities within the dense crowd. The errors in *x* and *y* coordinates are presented in Fig. 12.

There is a risk that increased activities, via random changes in z_l^c, will introduce increased uncertainty in the group tracking. Increased internal activities may be irrelevant to the actual crowd motion, but of interest from a crowd behavior point of view.

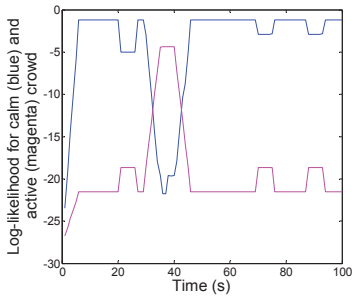

Figure 10. Log-likelihood for calm crowd (blue) and active crowd (mangenta). For 30 < *Time* < 40 the log-likelihood for active crowd is higher than for calm crowd.

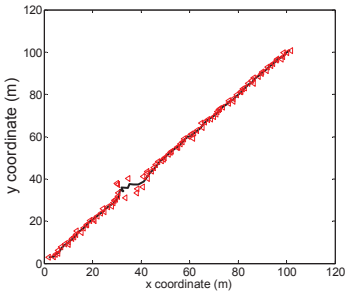

Figure 11. A crowd in motion. The black curve represents the estimated crowd position according to the group tracking and the red curve represents z_l^c.

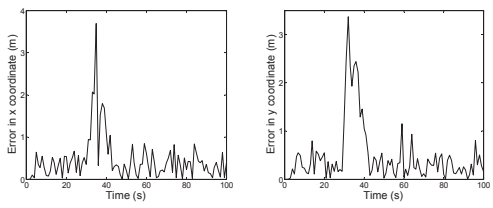

Figure 12. Position estimation error for the *x* and *y* coordinates. When the crowd is active during 30 < *Time* < 40 the position errors are increased considerably.

VI. Summary of Results and Conlcusions

This paper describes a preliminary framework for crowd analysis that can handle sparse and dense crowds. We have used *K*-means clustering to segment detections into groups and individuals. A dense crowd is equal to a dense cluster. In this paper the dense-cluster measure is adapted to a crowd size of approximately ten people. The dense-cluster measure can also be used to indicate what type of tracking that should be used, i.e. group tracking or MTT. When a dense crowd has been identified, group tracking should be used for the foreground detections in that area, instead of MTT. When the dense crowd has split, MTT should be used again.

In a sparse crowd the number of clusters equals the number of people in the scene. If people start to form dense clusters, the number of clusters will be reduced if no more people enter the scene. The number of cluster members will be increased for those clusters that add more people.

Dense-crowd behavior, for clusters with at least two members, is estimated using HMM. Input data are changes in the dense-cluster measure from frame to frame. For a cluster with only one member this method will probably not be the best approach. With only one member the behavior can be better estimated with detailed input data directly from MTT.

MTT can be performed with good accuracy as long as the crowd is sparse. In the first experiment we showed that people can still be detected in each frame in a dense crowd. However, track identities are continuously changed.

Table 1 summarizes the performance of the algorithms for experiments 1 to 3. The experiments show promising results with correct decisions ranging from 84 % and above.

Also some reflections on group tracking, and its relation to the dense crowd estimations, have been done. The results indicate that with increased crowd activity, group tracking performance will temporarily decrease. It is possible that group tracking accuracy can be improved if the knowledge on crowd behavior is used as feedback into the tracking process. However, this aspect of crowd analysis has to be investigated further before more final conclusions can be drawn.

The framework has been investigated for relatively small crowds. How will it work for large crowds, i.e. with hundreds of people? For large crowds, the dense-crowd measure and HMM parameters will be different. The new parameters should be obtained from training the algorithms on data sets representing large crowds. For example, fights between few people will not result in active crowd behavior if the crowd is very large.

TABLE I. Crowd Event Detection

Efficiency of the algorithms for split/merge and calm/active	**Experiments, type of event, length of scenario**			
	Ex 1 Merge/ Split (41 s)	***Ex 2 Merge/ Split (25 s)***	***Ex 2 Calm/ Active (25 s)***	***Ex 3 Calm/ Active (12 s)***
$\eta_{scenario}$ (%)	88	84	84	100

On the other hand, there may be other events that are of equal importance such as if people start to merge into a large and dense crowd, if the large and dense crowd is stationary or moves in some direction with a certain speed.

It is also possible that detections of persons should be based more directly on foreground pixels, instead of first detecting or classifying the foreground pixels as a person.

In conclusion, for smaller dense crowds of approximately ten people a combination of micro- and macroscopic approach seems to be a good basis for situation assessment in complicated environments (with for example occlusion). In sparse crowds individuals can be detected, tracked and recognized and consequently there is a possibility to collect detailed information of different events/actions in the scene. A dense crowd cannot give the same detailed information, but instead information on other types of events associated to the crowd as a whole.

In a real video surveillance system it would be advantageous to have some kind of combination of micro- and macroscopic approaches in order to reduce uncertainties in the automatic decisions and, at the same time, minimize the number of computations (as a result of lower demands on detections and a reduction in the number of tracks in dense environments).

VII. Future Work

The crowd analysis framework consists of several algorithms ranging from detection, tracking, group tracking and behavior analysis. Each algorithm has its own possibilities and limitations that can be further studied and developed. Moreover, the relations between the different algorithms can be further studied and developed in order to connect them properly in the framework. In the future work we plan to:

- Continue to develop the different parts of the framework and especially the relations between the different algorithms, i.e. between tracking (MTT and group tracking), clustering and behavior analysis.
- Evaluate the framework on larger data sets.
- Evaluate the framework for larger crowds.
- Investigate if information on crowd behavior can be used to enhance the group tracking performance.

References

[1] B. Zhan, D. N. Monekosso, P. Remagnino, S. A. Velastin, and L.-Q. Xu, "Crowd analysis: a survey," Machine Vision and Applications, vol. 19, pp. 354-357, 2008.

[2] D. Helbing, and P. Molnár, "Self-organization phenomena in pedestrian crowds", http://arxiv.org/abs/cond-mat/9806152v1, pp. 569-577, 1998.

[3] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," IEEE Int. Conf. on Computer, Vision and Pattern redcognition (CVPR), Miami, 2009.

[4] S. Zaidenberg, B. Boulay, C. Garate, D.-P. Chau, E. Corvée, and F. Brémond, "Group interaction and group tracking for video-surveillance in underground railway stations," International Workshop on Behaviour Analysis and Video Understanding (ICVS 2011), September, Sophia Antipolis, 2011.

[5] A. Gning, L. Mihaylova, S. Maskell, S. K. Pang, and S. Godsill, "Group object structure and state estimation with evolving networks and Monte Carlo Methods," IEEE Transactions on Signal Processing, vol. 59, no. 4, pp. 1383-1396, September, 2010.

[6] S. J. McKenna, S. Jabri, Z. Duric, H. Wechsler, and A. Rosenfeld, "Tracking groups of people," Computer Vision and Image Understanding, vol. 80, no. 1, pp. 42-56, 2002.

[7] K. Gilholm, S. Godsill, S. Maskell, and D. Salmond, "Poisson models for extended target and group tracking," SPIE Conference on Signal and Data Processing of Small Targets, August, San Diego, 2005.

[8] P. Skoglar, U. Orguner, D. Törnqvist, and F. Gustafsson, "Pedestrian tracking with infrared sensor using road network information," EURASIP Journal on Advances in Signal Processing, vol. 26, February, 2012.

[9] M. Andersson, J. Rydell, and J. Ahlberg, "Estimation of crowd behavior using sensor networks and sensor fusion," 12th Conf. on Information Fusion, Seattle, WA, pp. 396-403, July, 2009.

[10] M. Andersson, S. Ntalampiras, T. Ganchev, J. Rydell, J. Ahlberg, and N. Fakotakis, "Fusion of acoustic and optical sensor data for automatic fight detection in urban environments," 13th Conf. on Information Fusion, Edinburgh, July, 2010.

[11] S. Blackman, and R. Popoli, Design and Analysis of Modern Tracking Systems, Artech House, 1999.

[12] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference and Prediction, 2nd edition, Springer, 2009.

[13] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," In Proc. of the IEEE, vol. 77, no. 2, pp. 257-286, 1989.

[14] C. M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.

[15] www.ino.ca/Video-Analytics-Dataset